STUDIA UBB CHEMIA, LX, 4, 2015 (p. 213-228)
(RECOMMENDED CITATION)

STRUCTURAL MODELING OF PHENYLALANINE AMMONIA-LYASES AND RELATED MIO-CONTAINING ENZYMES – AN INSIGHT INTO THERMOSTABILITY AND IONIC INTERACTIONS

GERGELY BÁNÓCZI[a], CSONGOR SZABÓ[a], ZSÓFIA BATA[a,b], ALINA FILIP[c], GÁBOR HORNYÁNSZKY[a], LÁSZLÓ POPPE[a,d] *

ABSTRACT. Biocatalysis and bioinformatics are powerful tools to analyze and enhance the properties of biocatalysts for industrial technologies. In this work, the three-dimensional structures of various phenylalanine ammonia-lyases and other 5-methylene-3,5-dihydro-4*H*-imidazol-4-one (MIO) dependent enzymes were investigated, analyzed, and constructed. The three- dimensional structures of MIO-containing aromatic ammonia-lyases and 2,3-aminomutases from the PDB database were analyzed. In several instances the structural deficiencies and inactive conformations were modified to approach the active states. The checked / modified PDB structures of the MIO-enzymes served as templates for the large scale homology modeling for phenylalanine ammonia-lyase (PAL) sequences with unknown structure. Multiple settings were tested in the modeling workflow to result in structures with intact active center, good overall protein quality and reasonable number of salt bridges. The experimental temperature optima of the investigated PALs were correlated with two common factors of thermal stability: salt bridge and disulfide bridge numbers. Our study indicate *i*) a moderate correlation between salt bridge numbers and temperature optima, *ii*) and negligible effects of disulfide bridges. The two examples which do not correlate with the others indicate the presence of other important factors contributing to thermal stability of the MIO-enzymes. The modified PDB structures can serve further molecular modeling projects such as functional studies with a QM/MM approach.

Keywords: *3,5-dihydro-5-methylidene-4H-imidazol-4-one, MIO prosthetic group, homology modeling, phenylalanine ammonia-lyase, thermal stability, salt bridge*

[a] *Budapest University of Technology and Economics, Department of Organic Chemistry and Technology, Műegyetem rkp. 3., H-1111 Budapest, Hungary.*
[b] *Hungarian Academy of Sciences, Research Centre for Natural Sciences, Institute of Enzymology, Magyar tudósok körútja 2., H-1117 Budapest, Hungary.*
[c] *Babes-Bolyai University, Biocatalysis and Biotransformation Research Group, Str. Arany János nr. 11, RO-400028 Cluj-Napoca, Romania.*
[d] *SynBiocat LLC, Lázár deák u. 4/1., H-1173 Budapest, Hungary.*
* *Corresponding author: poppe@mail.bme.hu*

INTRODUCTION

As nowadays industrial development needs more efficient, yet green processes, new methods must meet more criteria than ever before. Biotechnology, one of the disciplines that were proven to successfully fulfill this task, can provide enzymes that are efficient and green catalysts. Enzymes, as biocatalysts, can be applied under mild conditions and at ambient temperature, can be highly selective, and can avoid the production of hazardous materials during their whole lifecycle. Bioinformatics can contribute to the experimental work required to select and develop the suitable biocatalyst by reducing the amount of required experimental work and increasing the amount of information extracted from the experiments.

An important feature of enzymes is their temperature tolerance. At elevated temperatures the reaction rates are higher and the probability of the biocatalyst digestion by microbial infection is lower. Generally, thermophile enzymes can tolerate heat treatment, as well as higher denaturing agent and substrate concentrations [1]. On the other hand, the biotechnological value of cold-adapted enzymes (aka psychrophiles) stems from their high catalytic activity at low to moderate temperatures providing energy savings to processes. Also, the use of psychrophiles can lower the risk of undesirable chemical reactions and can enable the transformation of thermally labile substrates [2].

The structural features of thermotolerance of enzymes have been intensively researched in the previous decades yielding a vast amount of data. Szilágyi *et al.* [3] concluded that different protein families adapt to higher temperatures by different sets of structural devices. Regarding the structural parameters, the only generally observed rule was an increase in the number of salt bridges with increasing growth temperature of host organisms. Other parameters showed just a trend, whereas the number of hydrogen bonds and the polarity of buried surfaces exhibited no clear-cut tendency to change with growth temperature. They suggested that proteins from extreme thermophiles are stabilized in different ways compared to the moderately thermophilic ones. The preferences of these two groups are different with regards to the number of ion pairs, the number of cavities, the polarity of exposed surface and the secondary structural composition.

Similarly, according to Scandurra *et al.* [1], stability of a protein from a thermophile, compared to its mesophilic homologue, is a property acquired through many small structural differences such as the modulation of the canonical forces, electrostatic (hydrogen bonds and ion-pairs) and hydrophobic interactions. Thermophilic proteins – produced by thermo- and hyperthermophilic microorganisms growing between 45 and 110 °C – show structural restriction on flexibility, which allows them to be functionally competent at elevated

temperatures, but renders them unusually rigid and less active at mesophilic temperatures (10-45 °C) due to increased compactness. From a comparative analysis of several families of proteins including at least one thermophilic structure in each family, it appeared that thermal stabilization was accompanied by an increase in hydrogen bonds and salt bridges [1]. Thermostability appeared to correlate also with a better packing within buried regions.

Kurnikova *et al.* [4] have determined flexible regions for four homologous pairs from thermophilic and mesophilic organisms by two computational methods: FoldUnfold [5] which is based on amino acid sequences, and MD/First [6] which use three-dimensional structures and molecular dynamics (MD). Both methods allowed determining flexible regions in protein structures. Supporting the previous statement by Scandurra *et al.* [1], MD simulations showed that thermophilic proteins were more rigid in comparison to their mesophilic homologues. It has been also found that the local networks of salt bridges and hydrogen bonds in thermophiles render their structure more stable with respect to fluctuations of individual contacts. This ionic network connects α-helices and rigidifies the structure. Mesophiles can be characterized by standalone salt bridges and hydrogen bonds or small ionic clusters. Such difference in the network of salt bridges results in different flexibility of homologous proteins. Despite these generalizations, no universal rules were found as structural factors leading to enhanced thermostability within a certain family of enzymes.

In contrast to their thermophilic counterparts, the structures of thermolabile, cold-adapted enzymes generally permit high flexibility [7,8]. According to Saunders *et al* [9] and Siddiqui *et al* [7], psychrophilic enzymes tend to possess various combinations of the following features: decreased core hydrophobicity, increased surface hydrophobicity, lower arginine/lysine ratio, more and longer loops, decreased secondary structure content, more glycine residues, less proline residues in loops and more proline residues in α-helices, a reduced number of disulfide bridges, fewer electrostatic interactions, reduced oligomerization, and an increase in conformational entropy of the unfolded state.

Aromatic ammonia-lyases catalyze the non-oxidative ammonia elimination from aromatic (*2S*)-amino acids, producing *α,β*-unsaturated carboxylic acids. The members of this enzyme family are the histidine ammonia-lyase (HAL, EC 4.3.1.3) [10,11], tyrosine ammonia-lyase (TAL, EC 4.3.1.23) [12], and phenylalanine ammonia-lyase (PAL, EC 4.3.1.24) [13] catalyzing the deamination of the corresponding (*2S*)-amino acids to (*E*)-cinnamic acid, (*E*)-urocanic acid, and (*E*)-coumaric acid, respectively (Figure 1/a). In plants, PAL catalyzes the first step of the phenylalanine degradation pathway, and thus the biosynthesis of several classes of phenylpropanoids, such as lignins, flavonoids, and coumarins [14]. The synthetic application of PAL is based on the stereoconstructive reverse reaction yielding (*2S*)-phenylalanine derivatives

from achiral precursors or the kinetic resolution of racemic amino acids providing access to (*2R*)-phenylalanine derivatives as residual substrates [15]. From an environmental point of view it is important that PAL lies at the branch of primary and secondary metabolism in plants and therefore it is a target for herbicides [16].

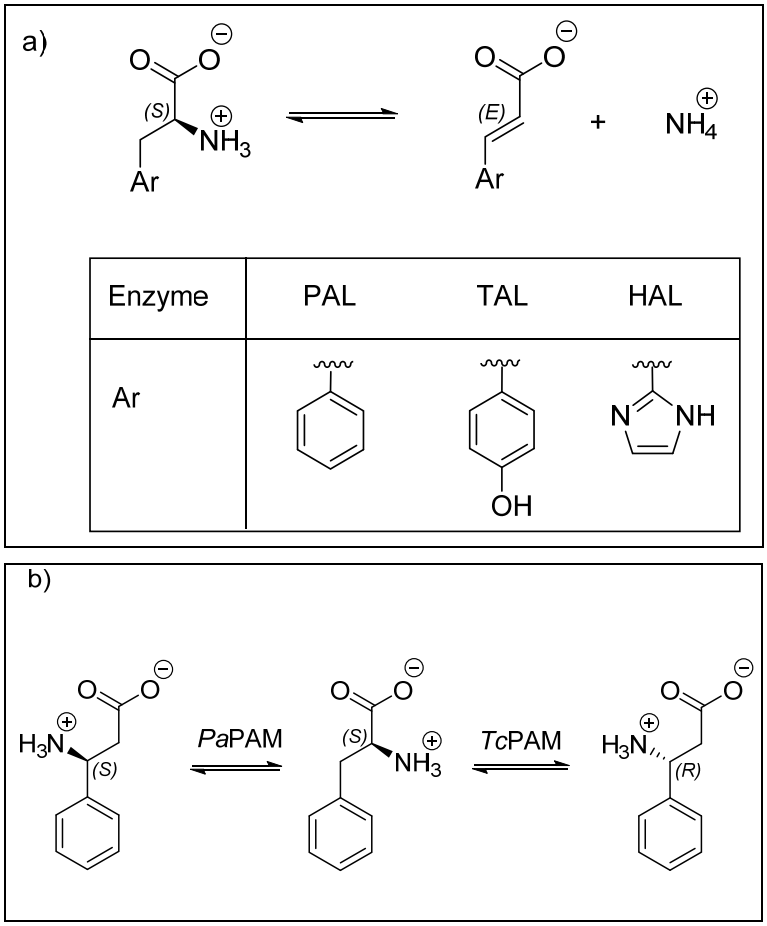

Figure 1. (a) reactions catalyzed by the aromatic ammonia-lyases and (b) phenylalanine 2,3-aminomutases.

The X-ray structure of HAL [17] revealed first that the enzyme contains a post-translationally formed 5-methylene-3,5-dihydro-4*H*-imidazol-4-one (MIO) as the electrophilic prosthetic group (Figure 2) [18]. This prosthetic group is formed autocatalytically by the elimination of two water molecules from an inner tripeptide Ala-Ser-Gly (in some cases Thr-Ser-Gly).

This prosthetic group can also be found in tyrosine 2,3-aminomutases (TAM, EC 5.4.3.6) and phenylalanine 2,3-aminomutases (PAM; (*3R*)-phenylalanine forming: EC 5.4.3.10; (*3S*)-phenylalanine forming: EC 5.4.3.11), which catalyze the selective formation of the 3-amino acids (Figure 1/b).

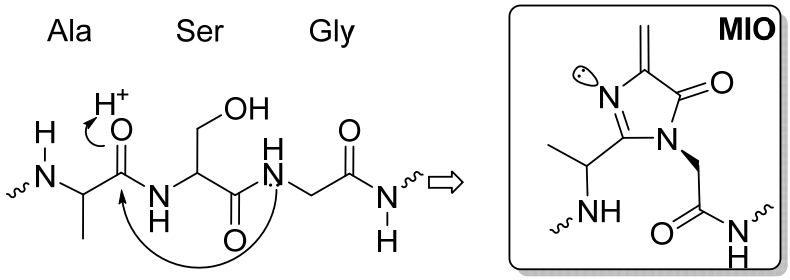

Figure 2. Formation of the 5-methylene-3,5-dihydro-4*H*-imidazol-4-one (MIO) prosthetic group from an Ala-Ser-Gly peptide fragment.

Only (3*S*)-tyrosine producing TAM structures have been determined so far, in contrary to PAMs, where both (3*R*)- and (3*S*)-phenylalanine producing structures are known (e. g. *Taxus canadensis* and *Pantoea agglomerans* PAM, respectively) [19].

In the Protein Data Bank (PDB) X-ray structures are available for enzymes from ten different organisms in which the MIO prosthetic group was identified: *Pseudomonas putida* HAL (PDB: 1GKM) [20], *Rhodosporidium toruloides* PAL (PDB: 1Y2M) [21], *Petroselinum crispum* PAL (PDB: 1W27) [22], *Anabaena variabilis* PAL (PDB: 3CZO) [23], *Nostoc punctiforme* PAL (PDB: 2NYF) [24], *Rhodobacter sphaeroides* TAL (PDB: 2O6Y) [25], *Streptomyces globisporus* TAM (PDB: 2QVE) [26], *Taxus canadensis* PAM (PDB: 3NZ4) [27], *Taxus chinensis* PAM (PDB: 2YII) [28], and *Pantoea agglomerans* PAM (PDB: 3UNV) [29].

In several cases, more than one structure is available, although only one representative is listed here for each MIO-enzyme.

An important structural difference between the previously listed enzymes is the presence of a ~170 amino acid long multihelix domain in the *C*-terminal region in enzymes with eukaryotic origin, against a ~7-19 amino acid long corresponding sequence segment in enzymes with prokaryotic origin (Figure 3/a). However, despite this considerable difference along with different substrates and function, the catalytic core domains of these proteins are remarkably similar in their tertiary and quaternary structures, deviating from a common fold considerably mainly in disordered, loop regions, and forming tetramers in every case (Figure 3/b).

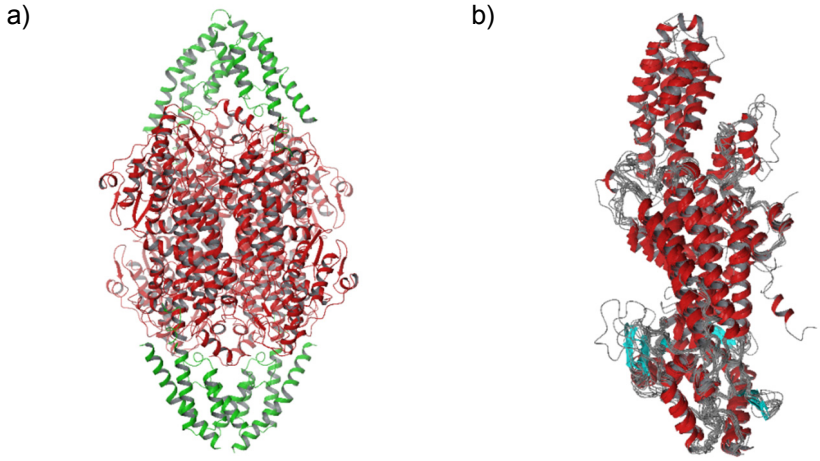

Figure 3. (a) Tetramer structure of *Tchin*PAM (PDB: 2YII). Ribbons in red indicate the catalytic core domain, common in MIO enzymes of both eukaryotic and prokaryotic origin. Ribbons in green represent the *C*-terminal multihelix domain, characteristic for enzymes of eukaryotic origin. (b) Overlay of the single chains of all MIO-enzymes listed in Table 1 indicated high structural conservation in the colored ordered secondary structure regions and less conservation in loop regions.

Mutagenesis studies proved the similarity of PAL [30] and HAL [31], and the importance of the Tyr110(PAL)/78(HAL) residue which was essential for the catalytic activity. This Tyr is situated in a flexible loop which is often missing from the crystal structures (PDB: 1Y2M) [21] or is in a supposedly inactive conformation (PDB: 1W27) [22]. However, in the structures of *Anabaena variabilis* PAL (PDB: 3CZO) [23], and *Rhodosporidium toruloides* TAL (PDB: 1Y2M) [21] compact active centers were found with tight Tyr-loop conformations, showcasing the supposedly active conformation.

EXPERIMENTAL

Experimentally determined protein structures from the Protein Data Bank (PDB) [32] were selected as representatives for each MIO-enzyme type (encoded by genes of different organisms, Table 1). Selection criteria were active site compactness, resolution, *R*-value, and *R*-free values. After the overlay of the single chains of the various MIO-enzymes, they were systematically checked for missing side chains, residues, and loops, for inactive Tyr-loop conformations and for large backbone deviations. Errors were corrected as much as possible by using partial homology modeling of the missing/erroneous part(s). The templates for homology modelling were chosen

based on the maximum similarity of the corresponding amino acid segments. The resulting checked/corrected experimental structures served as templates for the large scale homology modeling for those MIO-enzymes which did not have experimental structures.

In the *Pc*PAL structure the Tyr-loop and an adjacent loop were present in a catalytically inactive "loop-out" conformation [33], therefore these regions were reconstructed. Templates for the residues 98-131 and 324-350 were taken from *Tchi*PAM.

The Tyr-loop and a further loop which were both partly missing and deformed from the structure of *Rt*PAL crystal structure were modeled as well. The coordinates for the missing residues 103-124 and 343-358 of *Rt*PAL were taken from *Tchi*PAM as well.

In the *Pp*HAL and *Tcan*PAM structures the Tyr-loop is in a "loop-in" conformation [33], albeit in a partially open, non-active state. Residues 46-68 for *Pp*HAL and 68-88 for *Tcan*PAM were remodeled using the templates 2O6Y and 2YII, respectively.

Table 1. Experimental structures of aromatic ammonia-lyases and 2,3-aminomutases (occasionally with partial corrections) chosen as homology modeling templates

Organism	***Function***	***Abbreviation***	***PDB code***	***Uniprot code***	***Errors***	***Correction template***
Anabaena variabilis	PAL	*Av*PAL	3CZO	Q3M5Z3	-	-
Petroselinum crispum	PAL	*Pc*PAL	1W27	P24481	Inactive loop conformations	2YII
Rhodosporidium toruloides	PAL/TAL	*Rt*PAL	1Y2M	P11544	Missing Tyr-loop and residues	2YII
Pseudomonas putida	HAL	*Pp*HAL	1GKM	P21310	Non-tight Tyr-loop	2O6Y
Rhodobacter sphaeroides	TAL	*Rs*TAL	2O6Y	Q3IWB0	-	-
Streptomyces globisporus	TAM	*Sg*TAM	2QVE	Q8GMG0	-	-
Taxus canadensis	PAM	*Tcan*PAM	3NZ4	Q6GZO4	Non-tight Tyr-loop	2YII
Pantoea agglomerans	PAM	*Pa*PAM	3UNV	Q84FL5	-	-
Taxus chinensis	PAM	*Tchi*PAM	2YII	Q68G84	Missing residues	1W27

In the BRENDA enzyme database 13 different MIO-containing phenylalanine ammonia-lyases were found with known temperature optima but without experimentally determined structure [34] (Table 2). For each of the various PALs sixteen homology models of the catalytically active homotetrameric form were constructed each time, using MODELLER [35]. From the beginning

of the optimization process, the ASG amino acid triads in the conserved position of the MIO prosthetic group were replaced by the corresponding MIO-structures. Symmetry constraints were applied to the monomers forming the homotetramer.

Refinement of the raw models constructed by MODELLER was performed with Maestro [36] and Protein Preparation Wizard [37,38] (bond order assignment, addition of hydrogens). The refined models were minimized with TINKER [39] (forcefield: OPLS, minimization method: TNCG).

For the refinement of the initial models of *Rubrobacter xylanophylus* PAL [40], multiple settings were tested (max. 8 models for each) for MODELLER and for post-minimization with within TINKER. Varied settings were *i)* in MODELLER: the level of simulated annealing, all-atom/heavy atom-only modeling, monotemplate/multitemplate modeling, different sequence alignments, number of refinement stage iterations; *ii*) in TINKER post-minimization: solvent model (vacuum or implicit water) and convergence criterion (1.0, 0.1, and 0.05 kcal mol^{-1} $Å^{-1}$ RMS gradient).

Table 2. Phenylalanine ammonia-lyases with known temperature optima but without experimentally determined structures

Organism	*Isoform*	*Abbreviation*	*Uniprot code*	*Reference*
Arabidobsis thaliana	1	*At*PAL1	Q3M5Z3	41
Arabidobsis thaliana	2	*At*PAL2	P24481	41
Arabidobsis thaliana	3	*At*PAL3	Q7N4T3	41
Arabidobsis thaliana	4	*At*PAL4	P11544	41
Bambusa oldhamii	1	*Bo*PAL1	P21310	42
Bambusa oldhamii	2	*Bo*PAL2	Q7N296	42
Bambusa oldhamii	4	*Bo*PAL4	Q3IWB0	42
Cistanche deserticola	-	*Cd*PAL	Q8GMG0	43
Gossypium hirsutum	-	*Gh*PAL	Q6GZO4	44
Helianthus annuus	-	*Ha*PAL	Q84FL5	45
Jatropha curcas	-	*Jc*PAL	Q9KHJ9	46
Oryza sativa	-	*Os*PAL	A2X7F7	47
Ustilago maydis	-	*Um*PAL	Q96V77	48

The number of salt bridges were determined in the final models (which were evaluated by the active site geometry and overall protein health). For salt bridge statistics ANOVA, Mann-Whitney U-tests, and Wilcoxon matched-pair tests were applied, using Statistica [49]. The structures were validated with PROCHECK [50] using SAVES (Structure Analysis and Verification Server).

Salt bridge numbers were determined with VMD with default parameters [51]. Sulfur atoms of cysteine residues were considered as disulfide bridge candidates with an interatomic distance less than or equal to 4 Å, or less than or equal to 7 Å when at least one of the residues was located on a loop.

The final procedure consisted of multitemplate modeling followed by a thorough simulated annealing and minimization in vacuum to a RMS gradient threshold of 1.0 kcal mol^{-1} $Å^{-1}$ followed by further geometry optimization with implicit water solvation to a RMS gradient of 1.0 kcal mol^{-1} $Å^{-1}$.

RESULTS AND DISCUSSION

Our aim with this work was to investigate, analyze and construct the three-dimensional structures of various phenylalanine ammonia-lyases and further MIO-dependent enzymes. Partial and full homology models were created to check the possible correlations between common factors of thermostability and temperature optima (i.e. temperature of maximum reaction velocity).

Modeling in monomeric form followed by construction of the tetramer resulted in inappropriate, non-interacting adjacent loops and side chains, steric clashes. Furthermore, energy minimization of such tetrameric structures led to more distorted structures than that obtained by tetrameric homology modelling. This effect could be visualized in the case of partial homology modelling of the PAL1 from *Petroselinum crispum* (*Pc*PAL, see Figure 4). Therefore, the structures of the MIO-containing enzymes in this study were always modeled in tetrameric form.

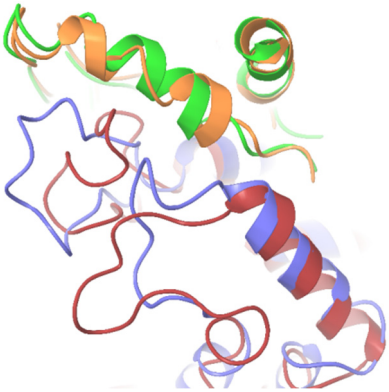

Figure 4. Comparison of the chains A and B of two different homology models of *Pc*PAL. Model 1 *(chain A: brown, chain B: orange*) was modeled in monomeric form which was assembled to tetramer, Model 2 *(chain A: blue, chain B: green)* was modeled as a tetramer from the beginning of the modelling process.

The tetrameric structures for *Pc*PAL were minimized to 0.1 kcal mol^{-1} $Å^{-1}$ RMS gradient in both cases using our standard procedure. Significant differences were found between the depicted loops of the two models. In the case of the tetramer built from monomeric model (Model 1), energy minimization resulted in a distortion of a helical part of chain B caused by the loop region of chain A highlighting the importance of modeling the homotetrameric structures as a complex (Model 2). Because the active site of the MIO-containing enzymes is formed by three different chains of the tetramer, such effects could not be neglected.

Although there is little consensus about the structural factors of thermostability, experts mainly agree on the importance of disulfide bridges and electrostatic interactions such as salt bridges, hydrogen bonds. Therefore, by using the modeled tetrameric structures of several phenylalanine ammonia-lyases, several modeling procedures resulting in high quality structures were analyzed with regards to active center compactness, overall protein health and reasonable salt bridge numbers. All tested settings (see experimental section) resulted in compact active centers, and structures with good protein health.

Unfortunately, because of the large deviations in salt bridge numbers among models within each test settings, all factors proved statistically insignificant (at $p<0.05$). Among the varied test settings, two factors showed remarkable effects on salt bridge numbers and protein health. The most significant changes were caused by the variations in the convergence criterion of post-minimization and by the use of implicit water model. Optimizations with the use of implicit water model resulted in better overall protein quality but significantly lower number of salt bridges. It is likely that the solvent model dampens the electrostatic attraction between ionized side chains preventing them from forming salt bridges. On the other hand, during extensive minimization with lower gradient thresholds, with or without solvent modeling, the protein quality deteriorated as the energy decreased, although the salt bridge numbers increased every time. This resulted in overestimation of salt bridge numbers.

Therefore, the most reasonable optimization protocol consisted of a pre-optimization of the models without solvent model to 1.0 kcal mol^{-1} $Å^{-1}$ RMS gradient followed by further optimization with solvent model to 1.0 kcal mol^{-1} $Å^{-1}$ RMS gradient again. Noteworthy that further optimization of the models to 0.05 kcal mol^{-1} $Å^{-1}$ RMS gradient resulted in higher salt bridge numbers, but also significantly worse protein health statistics. Also worth mentioning, that for the further optimized structures the salt bridge number statistics showed worse correlation with the experimental temperature optima values than in case of the less optimized (1.0 kcal mol^{-1} $Å^{-1}$ RMS gradient) structures. This indicates a possible over-refinement in the protocol including extensive minimization.

The average salt bridge numbers in the tetrameric models obtained in this way were correlated with the temperature optima (Table 3 and Figure 5).

Table 3. Salt bridge and disulfide bridge statistics of theoretical (as the average of numbers in the 16 models of each enzyme) and experimental PAL structures

Enzyme	*Salt bridge No. (Mean)*	*Salt bridge No. (RMSD)*	*Salt bridge fraction in the core region [%]*	*Salt bridge fraction in C-terminal domain [%]*	*Salt bridge fraction between the two regions [%]*	*Possible disulfide bridge number*	*Temp. optimum (°C)*
Eukaryotic PALs							
*At*PAL1	230	8,9	70.7	18.9	10.5	0	47
*At*PAL2	110	6.5	71.5	22.1	6.4	2	48
*At*PAL3	116	6.3	69.9	14.7	15.4	0	31
*At*PAL4	259	15.7	68.5	19.9	11.6	0	47
*Bo*PAL1	253	11.3	70.1	18.9	11.0	4	60
*Bo*PAL2	246	13.8	69.3	19.8	10.9	4	60
*Bo*PAL4	238	9.7	72.9	17.9	9.2	8	50
*Cd*PAL	253	12.7	68.1	21.5	10.4	4	55
*Gh*PAL	244	8.0	68.6	21.0	10.3	2	30
*Ha*PAL	214	8.7	67.1	21.8	11.1	0	55
*Jc*PAL	243	13.2	71.6	17.3	11.1	2	60
*Os*PAL	248	9.5	68.4	20.1	11.5	4	50
*Pc*PAL[a]	208	-	67.0	22.7	11.3	0	58
*Um*PAL	177	14.5	80.5	14.6	4.9	0	30
Prokaryotic PAL							
*Av*PAL[b]	74	-	94.6	5.4[c]	0	0	40

[a] Partial homology model [b] Experimental structure (PDB: 3CZO) [c] Fraction of saltbridges in the sequence segment corresponding to the *C*-terminal multihelix domain in enzymes with eukaryotic origin [%]

The MIO-containing PALs of eukaryotic origin in Table 3 could be split into two groups. When the correlation analysis was performed with inclusion of all points, almost no correlation between the salt bridge numbers and temperature optima were found (R^2=0.21). In contrary, a much better correlation (R^2=0.56) was found when outliers (*At*PAL2 and *Gh*PAL) representing a group defferent

from the main group were omitted from the analysis (Figure 5). The results for the main group, based on the comparison of salt bridge numbers from sophisticated homology models and experimental temperature optima of the modeled MIO-enzymes, implied reasonable correlation between salt bridges and temperature optima despite the heavily multivariate nature of the temperature optima of maximum reaction velocities. The two non-correlating enzymes in the other group indicated the importance of other factors which could also significantly contribute to the thermal dependence of the enzyme activity which may be influenced by local factors. Such local factors may be alterations of flexibility of the catalytically essential Tyr-containing loop.

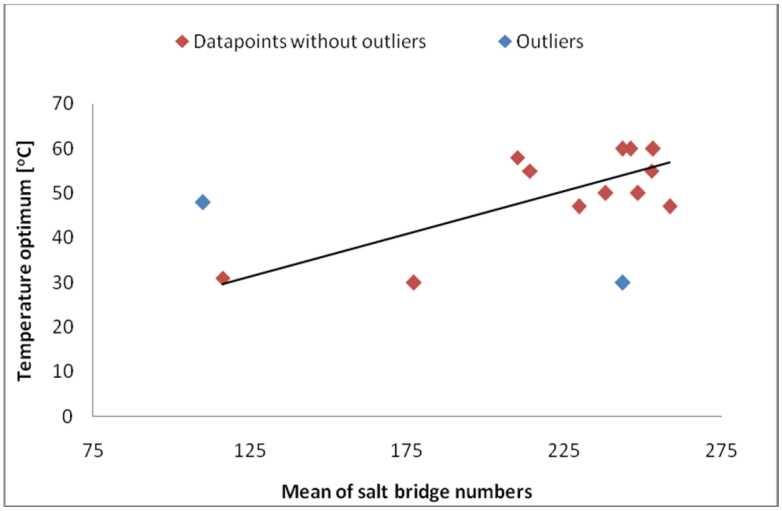

Figure 5. Graphical representation of the eukaryotic PAL salt bridge numbers (split into two groups) versus the temperature optima of maximum reaction velocities

Regarding enzymes of eukaryotic and prokaryotic origin, the salt bridge numbers were also decomposed, based the domain they belong to (Table 3). Surprisingly, there is no correlation between the calculated fractions and temperature optima, indicating that the *C*-terminal multihelix domain may not need additional ionic interactions, other than the overall increase in salt bridges, to stabilize.

The extremely low salt bridge number of *Av*PAL indicated that PALs of prokaryotic origin may need significantly less salt bridges to achieve a high temperature optima than those of eukaryotic origin. Although building a hypothesis upon only one data pair may be inadequate, this result would

further support the hypothesis on the Tyr-loop destabilizing / controlling role of the extended *C*-terminal helix domain in the eukaryotic PALs, which in not present in the prokaryotic ones [33]. To test the validity of this assumption, gathering more data in the future is required.

The number of possible disulfide bridges, however, showed no correlation with temperature optima independently of the method of modeling. One reason behind the absence of correlation between the disulfide bridges and temperature optima may be the lack of certain disulfide bridges despite the opportunity to be formed. This assumption lead to a warning: it could be possible that in several cases the actual recombinant host organisms were not the most appropriate ones to form all the possible disulfide bonds.

CONCLUSIONS

Our study indicated i) a moderate correlation between salt bridge number and temperature optimum, ii) and negligible effect of disulfide bridges with the temperature optima of phenylalanine ammonia-lyases. The two examples which do not correlate with the others indicate the presence of other important factors contributing to thermal dependence of activity for the MIO-enzymes, which may differ from the global thermostability. The modified PDB structures can support further molecular modeling projects such as functional studies by QM/MM approaches to identify and design better biocatalysts.

ACKNOWLEDGMENTS

We are thankful to Dr. Ödön Farkas for his remarks regarding this work. Financial support from OTKA (*NN-103242*) and from New Hungary Development Plan (*"Talent care and cultivation in the scientific workshops of BME" project, TÁMOP-4.2.2.B-10/1–2010–0009*) is acknowledged. *AF* thanks for the financial support of the Sectorial Operational Program for Human Resources Development 2007-2013, co-financed by the European Social Fund, under the project number POSDRU/159/1.5/S/132400.

REFERENCES

[1]. R. Scandurra, V. Consalvi, R. Chiaraluce, L. Politi, P.C. Engel, *Biochimie*, **1998**, *80*, 933.
[2.] R. Cavicchioli, T. Charlton, H. Ertan, S.M. Omar, K.S. Siddiqui, T.J. Williams, *Microbial Biotechnology,* **2011**, *4(4)*, 449.
[3]. A. Szilágyi, P. Závodszky, *Structure,* **2000**, *8,* 493.
[4]. T.B. Mamonova, A.V. Glyakina, O.V. Galzitskaya, M.G. Kurnikova, *Biochimica et Biophysica Acta,* **2013**, *1834*, 854.
[5]. O.V. Galzitskaya, S.O. Garbuzynskiy, M.Y. Lobanov, *Bioinformatics*, **2006**, *22, 2*948.
[6]. T. Mamonova, B. Hespenheide, R. Straub, M.F. Thorpe, M. Kurnikova, *Physical Biology*, **2005**, *2,* 137.
[7]. K.S. Siddiqui, R. Cavicchioli, *Annual Review of Biochemistry*, **2006**, *75,* 403.
[8]. G. Feller, „Protein Adaptation in Extremophiles", Nova Science Publishers, New York, **2008**, pp. 35–69.
[9]. N.F. Saunders, T. Thomas, P.M. Curmi, J.S. Mattick, E. Kuczek, R. Slade, *Genome Res*earch, **2003**, *13,* 1580.
[10]. I.L. Givot, T.A. Smith, R.H. Abeles, *Journal of Biological Chemistry*, **1969**, *244*, 6341.
[11]. R.B. Wickner, *Journal of Biological Chemistry*, **1969**, *244*, 6550.
[12]. J.A. Kyndt, T.E. Meyer, M.A. Cusanovich, J.J. Van Beeumen, *FEBS Letters*, **2002**, *512*, 240.
[13]. D.S. Hodgins, *Journal of Biological Chemistry*, **1971**, *246*, 2977.
[14]. K.R. Hanson, E.A. Havir, *Recent Advances in Phytochemistry,* **1978**, *12*, 91.
[15]. L. Poppe, J. Rétey, *Current Organic Chemistry*, **2003**, *7*, 1297.
[16]. L. Poppe, J. Rétey, *Angewandte Chemie, International Edition*, **2005**, *44*, 3668.
[17]. T.F. Schwede, J. Rétey, G.E. Schulz, *Biochemistry,* **1999**, *38(17)*, 5355.
[18]. L. Poppe, *Current Opinion in Chemical Biology,* **2001**, *5*, 512.
[19]. W. Szymanski, B. Wu, B. Weiner, S. Wildeman, B.L. Feringa, D.B. Janssen, *Journal of Organic Chemistry,* **2009**, *74*, 9152.
[20]. M. Baedeker, G.E. Schulz, *European Journal of Biochemistry*, **2002**, *269(6)*, 1790.
[21]. L. Wang, A. Gamez, C.N. Sarkissian, M. Straub, M.G. Patch, G.W. Han, S. Striepeke, P. Fitzpatrick, C.R. Scriver, R.C. Stevens, *Molecular Genetics and Metabolism,* **2005**, *86,* 134.
[22]. H. Ritter, G.E. Schulz, *Plant Cell*, **2004**, *16*, 3426.
[23]. L. Wang, A. Gamez, H. Archer, E.E. Abola, C.N. Sarkissian, P. Fitzpatrick, D. Wendt, Y. Zhang, M. Vellard, J. Bliesath, S.M. Bell, J.F. Lemontt, C.R. Scriver, R.C. Stevens, *Journal of Molecular Biology,* **2008**, *380*, 623.
[24]. M.C. Moffitt, G.V. Louie, M.E. Bowman, J. Pence, J.P. Noel, B.S. Moore, *Biochemistry*, **2007**, *46(4)*, 1004.

[25]. G.V. Louie, M.E. Bowman, M.C. Moffitt, T.J. Baiga, B.S. Moore, J.P. Noel, *Chemistry and Biology,* **2006**, *13*, 1327.
[26]. S.D. Christenson, W. Liu, M.D. Toney, B. Shen, *Journal of the American Chemical Society*, **2003**, *125(20)*, 6062.
[27]. L. Feng, U. Wanninayake, S. Strom, J. Geiger, W.D. Walker, *Biochemistry*, **2011**, *50(14)*, 2919.
[28]. B. Wu, W. Szymański, G.G. Wybenga, M.M. Heberling, S. Bartsch, S. de Wildeman, G.J. Poelarends, B.L. Feringa, B.W. Dijkstra, D.B. Janssen, *Angewandte Chemie International Edition,* **2012**, *51(2)*, 482.
[29]. S. Strom, U. Wanninayake, N.D. Ratnayake, K.D. Walker, J.H. Geiger, *Angewandte Chemie International Edition,* **2012**, *51(12),* 2898.
[30]. D. Röther, L. Poppe, S. Viergutz, B. Langer, J. Rétey, *European Journal of Biochemistry*, **2001**, *268,* 6011.
[31]. D. Röther, L. Poppe, G. Morlock, S. Viergutz, J. Rétey, *European Journal of Biochemistry*, **2002**, *269*, 3065.
[32]. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Research*, **2000**, *28*, 235.
[33]. S. Pilbák, A. Tomin, J. Rétey, L. Poppe, *The FEBS Journal,* **2006**, *273*, *1004.*
[34]. M. Scheer, A. Grote, A. Chang, I. Schomburg, C. Munaretto, M. Rother, C. Söhngen, M. Stelzer, J. Thiele, D. Schomburg, *Nucleic Acids Research*, **2011**, *39*, 670**.**
[35]. A. Sali, T.L. Blundell, *Journal of Molecular Biology*, **1993**, *234*, 779.
[36]. Maestro, version 9.7, Schrödinger, LLC, New York, NY, **2014**.
[37]. Protein Preparation Wizard 2014-1; Epik version 2.4, Schrödinger, LLC, New York, NY, 2014; Impact version 5.9, Schrödinger, LLC, New York, NY, 2014; Prime version 3.2, Schrödinger, LLC, New York, NY, 2014.
[38]. G.M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *Journal of Computer-Aided Molecular Design,* **2013**, *27*, 221.
[39]. TINKER Suite 7.1; *http://dasher.wustl.edu/tinker/*
[40]. K. Kovács, G. Bánóczi, A. Varga, I. Szabó, A. Holczinger, G. Hornyánszky, I. Zagyva, C. Paizs, G.B. Vértessy, L. Poppe, *PLoS ONE* **2014**, *9(1)*, e85943.
[41]. C.F. Cochrane, B.L. Davin, G.N. Lewis, *Phyochemistry*, **2004**, *65*, 1557.
[42]. L.S. Hsieh, G.J. Ma, C.C. Yang, P.D. Lee, *Phytochemistry,* **2010***, 71,* 1999.
[43]. S.G. Hu, M.J. Jia, J.Y. Hur, S.Y. Chung, H.J. Lee, J.D. Yuun, S.W. Chung, H.G. Yi, H.T. Kin, H.D. Kim, *Molecular Biology Reports*, **2011**, *38,* 3741.
[44]. A.J. Dubery, F. Smit, *Biochimica et Biophysica Acta,* **1994***, 1207,* 24.
[45]. J. Jorrín, R. López-Valbuena, M. Tena, *Biochimica et Biophysica Acta*, **1988**, *964*, 73.
[46]. J. Gao, S. Zhang, F. Cai, X. Zheng, N. Lin, X. Qin, Y. Ou, X. Gu, X. Zhu, Y. Xu, F. Chen, *Molecular Biology Reports*, **2012**, *39,* 3443.
[47]. D.A. Sarma, R. Sharma, *Phytochemistry*, **1999**, *50*, 729.
[48]. H.S. Kim, W.J. Kronstad, E.B. Ellis, *Phytochemistry*, **1996**, *43/2*, 351.
[49]. StatSoft, Inc. (2014). STATISTICA (data analysis software system), version 12. www.statsoft.com.

[50]. R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, J*ournal of Applied Crystallography*, **1993**, *26*, 283.
[51]. W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics*, **1996**, *14*, 33.